

Annual Report for Period:05/2010 - 04/2011**Submitted on:** 01/31/2011**Principal Investigator:** Loh, Gabriel H.**Award ID:** 0643500**Organization:** GA Tech Res Corp - GIT**Submitted By:**

Loh, Gabriel - Principal Investigator

Title:

CAREER: Computer Architecture Foundations for 3D-Integrated High-Performance Microprocessors

Project Participants**Senior Personnel****Name:** Loh, Gabriel**Worked for more than 160 Hours:** Yes**Contribution to Project:****Post-doc****Graduate Student****Name:** Puttaswamy, Kiran**Worked for more than 160 Hours:** Yes**Contribution to Project:**

PhD student conducting thesis research on the topic funded by this award. The student graduated this past calendar year and has taken a position with Intel Corporation.

Name: Kron, Jon**Worked for more than 160 Hours:** Yes**Contribution to Project:**

This is a new student who has performed some initial exploratory work on 3D implementations of the UT-Austin TRIPS processor, and is currently building tools to assist in the visualization of 3D processor performance issues.

Name: Shen, Guanhao**Worked for more than 160 Hours:** No**Contribution to Project:**

New student working on 3D-stacked DRAM cache architectures, but supported through a different program.

Undergraduate Student**Technician, Programmer****Other Participant****Research Experience for Undergraduates****Organizational Partners****Northwestern University**

Prof. Gokhan Memik and the PI (Loh) collaborated on a project that serendipitously arose from unrelated interactions (Prof. Memik had requested the PI to read a section of his CAREER proposal that discussed 3D integration, which in turn lead to other discussions and ultimately a submission to ISCA2008). This work is described in more detail elsewhere in this annual report.

Other Collaborators or Contacts

Prof. Gokahn Memik, Northwestern University.

Prof. Yuan Xie, Penn State University.

Activities and Findings

Research and Education Activities:

1. Exploiting 3D integration to reduce the impact of parametric variations and to improve yield (joint work with Prof. Gokhan Memik at Northwestern University).
2. Exploration and design of memory architectures involving multiple layers of 3D-stacked DRAM on top of multi-core processors.
3. Exploration and design of cache organizations implemented with small amounts of 3D-stacked DRAM.

[See FINDINGS section for technical summaries.]

4. Implementation of special-topics graduate course on advanced issues in processor microarchitectures. The course incorporates treatment of design for performance, as well as the many physical issues (power, thermals, area, cost) that are becoming increasingly important in 2D processors and especially in 3D processors. This course was offered during the Spring 2008 semester. Since there are no appropriate textbooks for this type of course, many new slides were created which will be made freely available for other professors and teachers to make use of (anticipated release of April 2009).
5. X86 Microprocessor simulator developed in support of part of this project has been publicly released as well as used in the graduate course listed above.

Findings:

1. Exploiting 3D integration to reduce the impact of parametric variations and to improve yield. This is joint work with Prof. Gokhan Memik at Northwestern University. In this work, we partition circuit paths over multiple circuit layers to ameliorate the effects of local/spatially-correlated parametric variations. By incorporating device from multiple layers, we produce an 'averaging' effect which improves yield and batch performance by reducing the probability of a circuit's path being systematically biased in one direction (e.g., all transistors in a circuit being slow). Our results indicate that a split architecture achieves a 36.9% lower yield loss rate and the average performance of the manufactured chips is increased by 15.3% compared to the same pipeline implemented on a 2D architecture. (Presented at DAC 2010)
2. While previous studies have proposed stacking a system's entire main memory directly on top of the processor with 3D integration, these proposed architectures are still limited by traditional CPU-memory interfaces that grossly underutilize the available bandwidth

provided by 3D. We evaluated different interface options that specifically avoid the conventional limitations of non-3D interfaces to greatly increase the amount of memory-request-level parallelism utilized by the overall system. To further improve performance, we also revisit the organization of the processor's miss handling architecture to further increase the number of cache misses that the processor can expose to the 3D-stacked memory. Our optimizations deliver over 100% more performance on top of the benefits of a naive 3D stacking of main memory. (Presented at ISCA 2008)

3. Many previous studies have explored stacking a single layer of DRAM on top of a processor to serve as a large last level cache (L2 or L3). Commodity DRAM is preferred for economic reasons, but the structure of commodity DRAM have differences from traditional SRAM-based cache implementations. Many past works have shown that there still exists a large performance gap between conventional cache replacement policies such as LRU and similar approximations thereof and optimal replacement decisions. We are currently exploring the opportunities to improve cache replacement and overall cache utilization in very large DRAM-based 3D-stacked caches. The DRAM row buffers are very wide, and present an opportunity to implement very highly associative cache structures. (Presented at MICRO 2009)

Training and Development:

Jonathan Kron (MS student): This project has provided the student with a variety of implementation and analysis experiences. The student has worked on multiple 3D projects with very challenging software infrastructures to tackle. The student unfortunately decided to pursue a high-paying job with an investment bank on Wall Street.

Guanhao Shen (PhD student): This PhD student is currently heavily involved in the implementation of a memory controller for 3D-stacked DRAM memory. He is also researching how to extend the design and organization of multiple memory controllers interfacing with a 3D-stacked memory to efficiently support future multi-/many-core processors where the number of cores is much greater than the number of memory controllers.

Outreach Activities:

A very well attended 3D tutorial was offered at ISCA 2008 in Beijing which included a strong participation from the computer architecture community as well as many Chinese student participants who are not usually able to attend our conferences. This was a collaborative effort with Prof. Yuan Xie (Penn State) and included industry speakers from Intel and IBM.

Research presentations/talk on the research supported by this grant have been given at IBM, AMD, Samsung, UT Austin, NCSU, Northwestern, UBC, Korea University.

Journal Publications

Loh, GH, "3D-Stacked memory architectures for multi-core processors", ISCA 2008 PROCEEDINGS: 35TH INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE, p. 453, vol. , (2008). Published,

Kiran Puttaswamy, Gabriel H. Loh, "3D-integrated SRAM Components for High-Performance Microprocessors", IEEE Transactions on Computers, p. , vol. , (2008). Accepted,

Gabriel H. Loh, "A Modular 3D Processor for Flexible Product Design and Technology Migrations", ACM Computing Frontiers Conference, p. , vol. , (2008). Published,

Gabriel H. Loh, "Extending the Effectiveness of 3D-Stacked DRAM Caches with an Adaptive Multi-Queue Policy", ACM/IEEE International Symposium on Microarchitecture, p. , vol. , (2009). Published,

Serkan Ozdemir, Pan Yan, Abhishek Das, Gabriel H. Loh, Gokhan Memik, Alok Choudhary, "Quantifying and Coping with Parametric Variations in 3D-Stacked Microarchitectures", ACM Design Automation Conference, p. , vol. , (2010). Published,

Books or Other One-time Publications

Gabriel Loh, "3D Microprocessor Design (Chapter)", (2009). Book, Accepted

Editor(s): Jason Cong, Sachin Sapatnekar and Yuan Xie

Collection: Three Dimensional Integrated Circuits Design: EDA, Design and Microarchitecture

Bibliography: Springer

Web/Internet Site

Other Specific Products

Contributions

Contributions within Discipline:

1. Exploiting 3D integration to deal with parametric variations: this is the first work (to our knowledge) that explicitly deals with parametric-variation-induced yield problems in 3D processors at the microarchitectural level.
2. 3D DRAM Caches: this work reconsiders the use of 3D-stacked DRAMs as large on-chip caches, and exploits the physical structures of DRAMs to enable new cache management optimizations that target a variety of behaviors due to different memory access locality patterns as well as the effects of multiple cores/threads contending for the cache resources. The results demonstrate by reconsidering conventional/straight-forward applications of 3D technology, there remains significant additional performance to be gained.
3. Re-examining 3D DRAM architectures: this work extends previous works by examining a much broader range of potential 3D DRAM designs, and the simulation results show that the performance benefits of our optimized 3D-stacked DRAMs are significantly greater than previously reported.
3. Modular 3D microarchitectures: this work extends the 'snap-on' 3D concept (such as that proposed by Mysore et al. in ASPLOS 2006) to provide non-invasive 3D-stacked performance enhancements. Previous snap-on proposals have only targeted relatively small, specific markets (e.g., debuggers/system developers) whereas this is the first work to our knowledge that leverages the 3D snap-on concept to deliver value to a much wider market (e.g., servers, workstations and extreme gaming platforms).

Contributions to Other Disciplines:

Contributions to Human Resource Development:

Graduation of Kiran Puttaswamy (PhD), who is now working for Intel Corporation in Austin, TX.

Graduation of Jonathan Kron (MS), who is now working on Wall Street.

Contributions to Resources for Research and Education:

A derivative of this work is a highly detailed microarchitecture simulation infrastructure partially derived from the SimpleScalar/x86 toolset (pre-release version). This simulator has been deployed in a graduate-level advanced topics course (from which two students published peer-reviewed workshops based on their project work), and has been made publicly (and freely) available for the community to use to aid in research and education. The corresponding lecture materials from the course have also been made available online. These resources can now be found online at zesto.cc.gatech.edu.

Based partially on work funded by this grant, a book chapter entitled '3D Microprocessor Design' has been accepted for publication in Cong, Sapatnekar and Xie's textbook titled 'Three Dimensional Integrated Circuits Design: EDA, Design and Microarchitectures' to be published by Springer.

Contributions Beyond Science and Engineering:

Conference Proceedings

Special Requirements

Special reporting requirements: None

Change in Objectives or Scope: None

Animal, Human Subjects, Biohazards: None

Categories for which nothing is reported:

Any Web/Internet Site

Any Product

Contributions: To Any Other Disciplines

Contributions: To Any Beyond Science and Engineering

Any Conference